

Proyecto Calidad del Aire

Observaciones de la calidad del aire en Valencia
desde varios puntos de observación



Índice



Nuestras Fuentes de Datos para el Proyecto

Para comprender y modelar la calidad del aire en Valencia, hemos integrado un conjunto diverso de fuentes de datos, capturando información en tiempo real, histórica y contextual

Cámaras

Se capturan y analizan imágenes de tráfico en Valencia para identificar patrones de movilidad y su relación con los niveles de contaminación. Esta información visual permite comprender mejor cómo influyen las dinámicas urbanas en la calidad del aire.

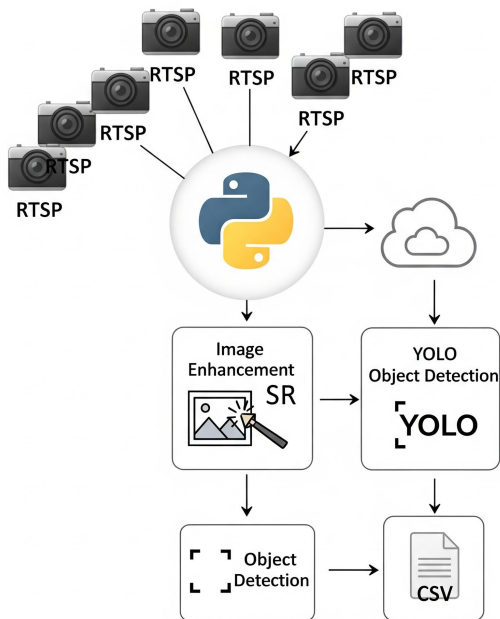
APIs

Los datos ambientales se obtienen en tiempo real mediante APIs, que proporcionan mediciones horarias clave (como niveles de NO₂, PM10, entre otros). Estas fuentes automatizadas permiten un monitoreo continuo y preciso de la calidad del aire.


PDFs


La documentación científica y técnica en formato PDF se utiliza para contextualizar y validar los resultados del proyecto. Estos documentos aportan referencias sólidas que respaldan el análisis y fortalecen las conclusiones obtenidas.


1. Imágenes de Tráfico en Tiempo Real




 **Fuente:** Cámaras de tráfico de la ciudad de Valencia.

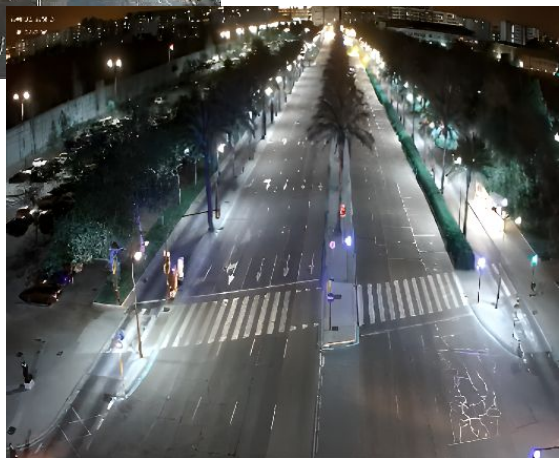
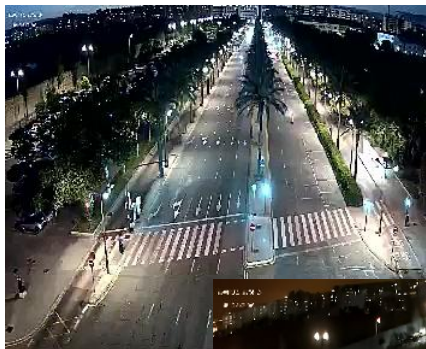
 **Datos obtenidos:** Capturas constantes de secuencias de imágenes, representando el flujo vehicular en distintas zonas urbanas.

 **Método de recopilación:** Scripts desarrollados en Python se conectan a las cámaras mediante el protocolo RTSP (Real-Time Streaming Protocol), realizando capturas periódicas de manera automatizada.

 **Almacenamiento:** Las imágenes recopiladas se almacenan de forma continua en la nube, específicamente en Azure Blob Storage, asegurando disponibilidad y escalabilidad.

 **Propósito:** Estas imágenes permiten realizar un análisis visual del nivel de congestión vial, identificar patrones de tráfico en diferentes franjas horarias y zonas, y estudiar su relación con los niveles de contaminación atmosférica registrados.

Mejora de Imágenes



Real-ESRGAN es un **modelo de IA de super-resolución basado en redes neuronales generativas adversariales (GAN)**, específicamente diseñado para la restauración práctica de imágenes en el mundo real.

Características principales:

Funcionalidad: Real-ESRGAN tiene como objetivo desarrollar algoritmos prácticos para la restauración general de imágenes/video. Su objetivo principal es mejorar la resolución y calidad de imágenes.

- Compresión JPEG
- Ruido y desenfoque
- Imágenes de baja resolución
-

Capacidades técnicas:

- Ampliación de imágenes hasta 4x (y escalas arbitrarias)
- Restauración de rostros cuando se integra con GFPGAN
- Soporte para imágenes con canal alfa, escala de grises y 16-bit
- Procesamiento por mosaicos para imágenes grandes

Es especialmente útil para mejorar fotografías antiguas, imágenes de baja calidad, capturas de pantalla comprimidas y arte digital, ofreciendo resultados más realistas que los métodos tradicionales de interpolación.



Yolo (You Only Look Once)



¿Qué es?

Algoritmo de IA que detecta y clasifica múltiples objetos en una imagen con **una sola pasada** de red neuronal

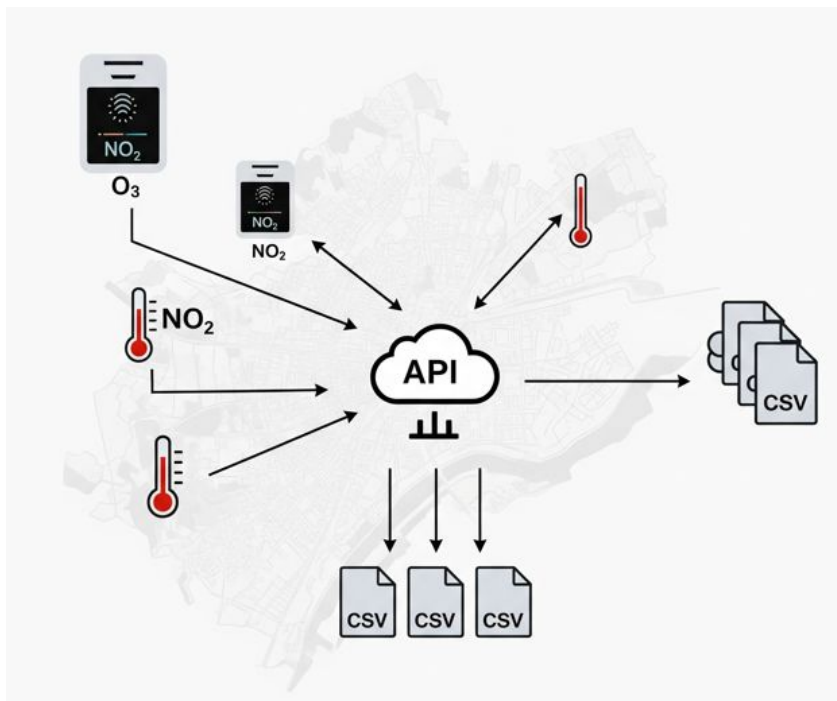
¿Cómo funciona?

⚡ **Divide** la imagen en cuadrícula → **Predice** cajas delimitadoras + confianza + clasificación **simultáneamente**

Ventajas Clave

- ✓ **Velocidad:** 30+ FPS en tiempo real
- ✓ **Eficiencia:** Una red para todo el proceso
- ✓ **Precisión:** Detecta múltiples objetos con alta exactitud

2. Datos de Sensores Ambientales (APIs)



Fuentes:

- API de Temperatura: Datos de sensores distribuidos en distintos puntos de Valencia.
- API de Calidad del Aire: Información proporcionada por la red oficial de estaciones de monitoreo ambiental.

Datos obtenidos:

Mediciones horarias de temperatura, niveles de contaminantes atmosféricos (NO₂, SO₂, O₃, PM10, PM2.5, entre otros) y variables meteorológicas relevantes como humedad o velocidad del viento.

Método de recopilación:

Scripts automatizados en Python acceden a las APIs cada hora, extrayendo los datos de manera programada para garantizar consistencia y actualización continua.

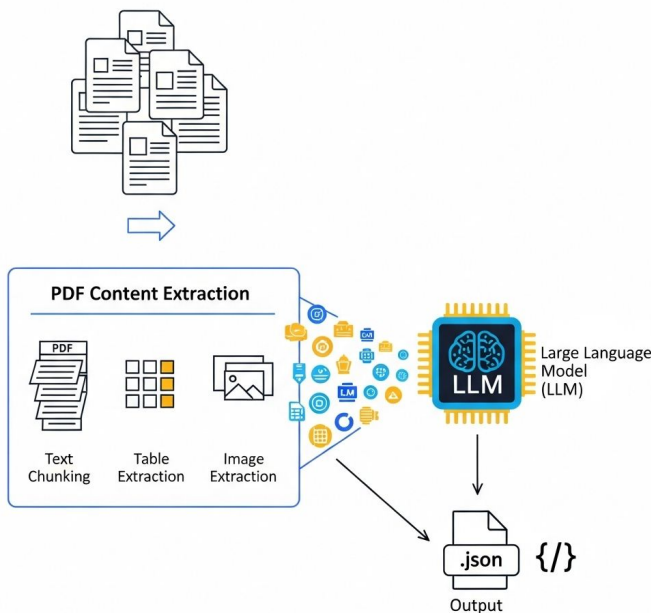
Almacenamiento:

Los datos se estructuran en archivos CSV por estación y tipo de variable, que se almacenan en la nube, permitiendo una gestión eficiente y un acceso rápido para análisis posteriores.

Propósito:

Capturar información cuantitativa y continua de las principales variables que influyen en la calidad del aire, permitiendo correlaciones con factores como el tráfico o las condiciones meteorológicas.

3. Documentación Científica y Técnica (PDFs)



Fuentes:

- Publicaciones científicas revisadas por pares.
- Informes técnicos de organismos especializados.
- Documentos oficiales de entidades gubernamentales como el Ayuntamiento de Valencia, la Generalitat Valenciana, el Gobierno de España y agencias europeas.

Información obtenida:

Estudios previos sobre calidad del aire, metodologías de análisis validadas, normativas medioambientales vigentes, planes de acción locales y europeos, datos históricos consolidados y factores clave ya identificados por expertos.

Método de recopilación:

Revisión manual y descarga desde repositorios académicos, bases de datos técnicas y portales institucionales oficiales.

Almacenamiento:

Toda la documentación se organiza y almacena en el servidor Elasticsearch, lo que facilita su consulta estructurada por tema y fuente.

Propósito:

Aportar un marco científico y regulatorio sólido que sirva para validar los hallazgos del proyecto, contextualizar los datos empíricos y alinear el análisis con las políticas e iniciativas actuales en materia de calidad del aire.



Integración de Datos con Azure DataFactory

Los flujos de datos en Azure DataFactory permiten integrar y procesar múltiples fuentes heterogéneas (aire, temperatura y cámaras), generando tablas raw organizadas y limpias que son almacenadas en Databricks para su análisis posterior.



● Flujo de Aire

Agrega los datos procedentes de **11 sensores de calidad del aire** distribuidos en distintos puntos de Valencia. Este flujo consolida las mediciones horarias de contaminantes en una única tabla estructurada.

● Flujo de Temperatura

Integra la información de **5 estaciones meteorológicas oficiales**, unificando variables como temperatura, humedad y presión atmosférica, esenciales para contextualizar los niveles de contaminación.

● Flujo de Cámaras

Procesa las imágenes capturadas por las cámaras de tráfico, **asignando un identificador único a cada imagen**. Esto permite evitar duplicados en los CSV finales y mantener la integridad de los datos visuales.

Parámetros

Variables

Configuración

Salida

Id. de ejecución de canalización

783450e9-bd15-4677-ac19-778f9c340a73

Estado de la canalización

Correcto

All status

Supervisión en métricas de Azure

Exportar a CSV

Mostrando elementos del 1 al 3 de un total de 3

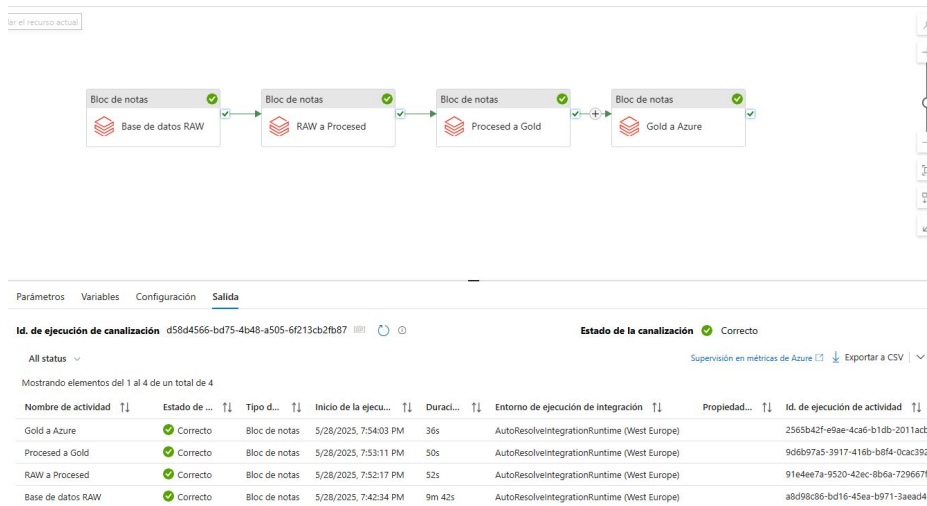
Nombre de actividad	Estado de ...	Tipo d...	Inicio de la ejecu...	Duraci...	Entorno de ejecución de integración	Propiedad...	Id. de ejecución de actividad
Camaras	<div></div> Correcto	Flujo de datos	5/28/2025, 7:09:42 PM	49s	AutoResolveIntegrationRuntime (West Europe)		ba57ccf1-ac1b-46f5-b63a-e102828f...
Temperatura	<div></div> Correcto	Flujo de datos	5/28/2025, 7:08:14 PM	1m 27s	AutoResolveIntegrationRuntime (West Europe)		69f867ef-4caf-4274-83f4-b9ee99eba...
Aire	<div></div> Correcto	Flujo de datos	5/28/2025, 7:03:31 PM	4m 41s	AutoResolveIntegrationRuntime (West Europe)		9cf09e5d-37d1-4c59-bf45-3ece24a1...



PipeLine DataFactory – Databricks

Este punto trata de optimizar la calidad, estructura y dar valor analítico de los datos para su explotación eficiente, manteniendo una arquitectura clara tipo lakehouse

(Raw > Processed > Gold).



Entrada:

Se parte de los archivos CSV almacenados en Azure Blob Storage, generados por la primera fase del flujo en DataFactory.

Proceso de Transformación en 4 Fases:

- 1. RAW (Bronce):**
Se cargan directamente los CSVs tal como están, sin ninguna modificación. Esta capa actúa como respaldo integro y punto de trazabilidad de los datos originales.
- 2. Processed (Plata):**
Se realiza limpieza y transformación inicial:
 - Conversión de tipos de datos mediante cast (por ejemplo: fechas, números, booleanos).
 - Eliminación de registros corruptos o incompletos.
 - Estandarización de columnas y formatos.
- 3. Gold (Oro):**
Se construyen tablas enriquecidas y resumidas, con lógica de negocio añadida:
 - Cálculo de métricas derivadas.
 - Inclusión de nuevas variables como si es fin de semana, franja horaria (día/noche), etc.
 - Agregaciones para análisis más rápidos.
- 4. Exportación a Azure:**
Las tablas Gold se exportan nuevamente como archivos CSV a Azure Blob Storage, listas para ser utilizadas por aplicaciones, dashboards o modelos de análisis.

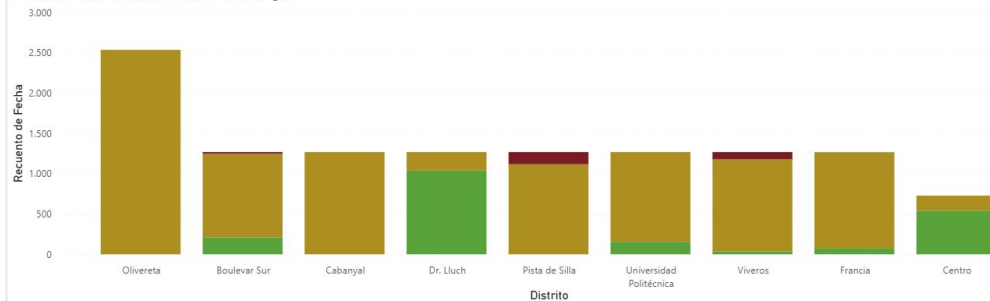
GOLD DATA - Power BI - Aire

Dashboard de análisis de contaminantes y calidad del aire

Districto	Mes	Día	Districto	tipozona	co	o3	pm10	pm25	so2	no2
Todas	mayo	18	Francia	Urbana	0	580	260	180	20	190
Calidad Aire	mayo	18	Francia	Urbana	0	630	260	180	20	170
Todas	mayo	18	Francia	Urbana	0	670	260	180	20	50
	mayo	18	Francia	Urbana	0	710	260	180	20	60
	mayo	18	Francia	Urbana	0	760	260	180	20	50
	mayo	18	Francia	Urbana	0	830	260	180	20	70
	mayo	18	Francia	Urbana	0	860	260	180	20	50
	mayo	18	Francia	Urbana	0	860	260	180	20	70
	mayo	19	Francia	Urbana	0	870	260	180	20	40
	mayo	19	Francia	Urbana	0	880	270	180	10	50
	mayo	19	Francia	Urbana	0	880	270	180	20	60

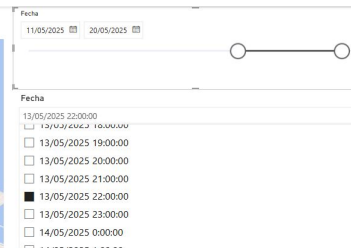
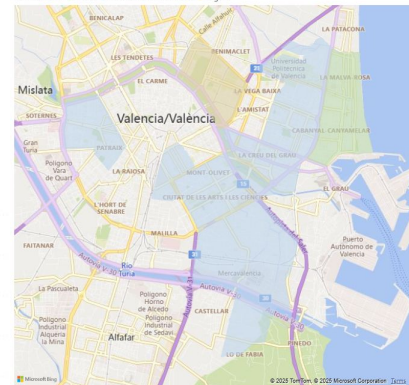
Recuento de Fecha por Distrito y Calidad Aire

Calidad Aire ● Buena ● Razonablemente Buena ● Regular



Código Postal y Calidad Aire

Calidad Aire ● Buena ● Razonablemente Buena ● Regular



GOLD DATA - Power BI - Temperatura

Dashboard de análisis de contaminantes y calidad del aire

fecha_sarg 24/04/2025 22/05/2025

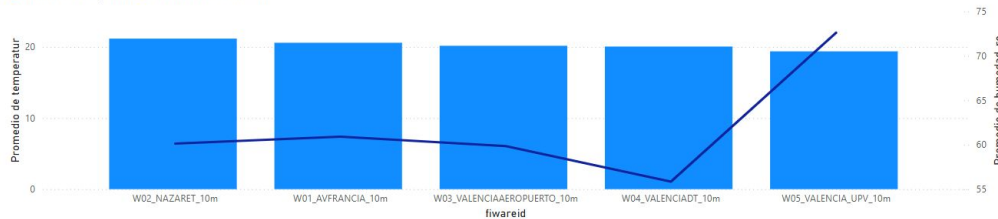
fiwareid

- ☐ W01_AVFRANCIA_10m
- ☐ W02_NAZARET_10m
- ☐ W03_VALENCIAAEROPUERTO_10m
- ☐ W04_VALENCIAIADT_10m
- ☐ W05_VALENCIA_UPV_10m

fiwareid	Mín. de temperatur	Máx. de temperatur	Suma de precipitac	Promedio de viento_vel	Mes	Día
W01_AVFRANCIA_10m	22	22	0	8	abril	24
W02_NAZARET_10m	22	22	0	4	abril	24
W03_VALENCIAAEROPUERTO_10m	24	24	0	6	abril	24
W04_VALENCIAIADT_10m	22	22	0	3	abril	24
W05_VALENCIA_UPV_10m	20	21	0	1	abril	24
W01_AVFRANCIA_10m	17	21	0	2	abril	25
W02_NAZARET_10m	18	22	0	2	abril	25
W03_VALENCIAAEROPUERTO_10m	17	22	0	4	abril	25
W04_VALENCIAIADT_10m	17	21	0	1	abril	25
W05_VALENCIA_UPV_10m	16	19	0	1	abril	25
W01_AVFRANCIA_10m	16	23	0	2	abril	26
W02_NAZARET_10m	16	23	0	1	abril	26
W03_VALENCIAAEROPUERTO_10m	14	23	0	3	abril	26
W04_VALENCIAIADT_10m	15	22	0	1	abril	26
W05_VALENCIA_UPV_10m	15	21	0	0	abril	26
W01_AVFRANCIA_10m	16	23	0	3	abril	27
W02_NAZARET_10m	16	24	0	3	abril	27
Total	11	32	89	2		

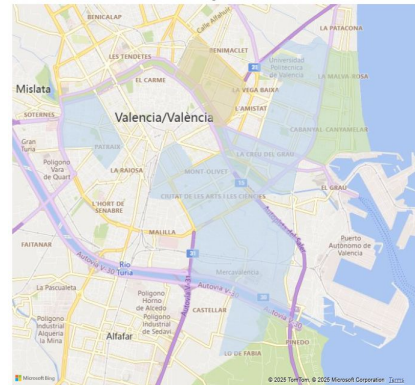
Promedio de temperatur y Promedio de humedad_re por fiwareid

● Promedio de temperatur ● Promedio de humedad_re



CodigoPostal y Calidad Aire

Calidad Aire ● Buena ● Razonablemente Buena ● Regular



Fecha

11/05/2025 20/05/2025

Fecha

13/05/2025 22:00:00

13/05/2025 19:00:00

13/05/2025 20:00:00

13/05/2025 21:00:00

13/05/2025 22:00:00

13/05/2025 23:00:00

14/05/2025 00:00:00

14/05/2025 01:00:00

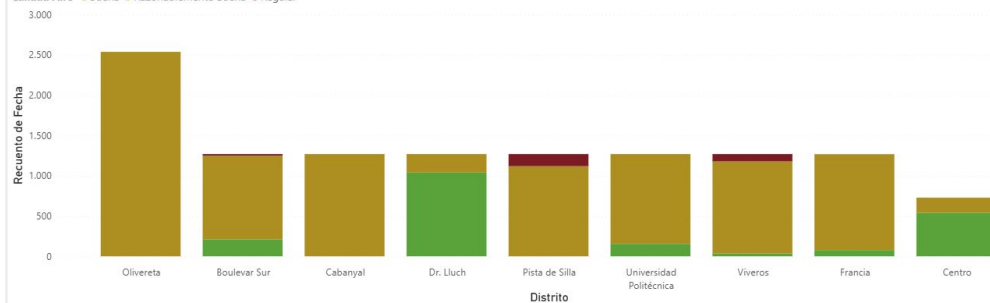
GOLD DATA - Power BI - Cámaras

Dashboard de análisis de contaminantes y calidad del aire

Districto	Mes	Día	Districto	tipozona	co	o3	pm10	pm25	so2	no2
Todas	mayo	18	Francia	Urbana	0	580	260	180	20	190
Calidad Aire	mayo	18	Francia	Urbana	0	630	260	180	20	170
Todas	mayo	18	Francia	Urbana	0	670	260	180	20	50
	mayo	18	Francia	Urbana	0	710	260	180	20	60
	mayo	18	Francia	Urbana	0	760	260	180	20	50
	mayo	18	Francia	Urbana	0	830	260	180	20	70
	mayo	18	Francia	Urbana	0	860	260	180	20	50
	mayo	18	Francia	Urbana	0	860	260	180	20	70
	mayo	19	Francia	Urbana	0	870	260	180	20	40
	mayo	19	Francia	Urbana	0	880	270	180	10	50
	mayo	19	Francia	Urbana	0	880	270	180	20	60

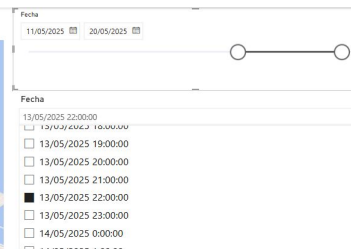
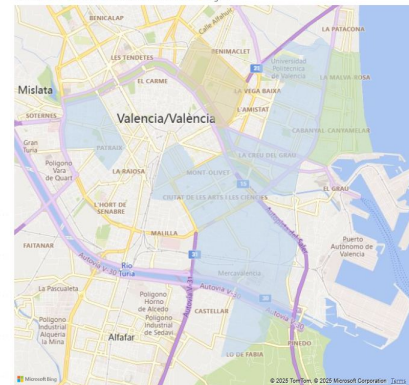
Recuento de Fecha por Distrito y Calidad Aire

Calidad Aire ● Buena ● Razonablemente Buena ● Regular



Código Postal y Calidad Aire

Calidad Aire ● Buena ● Razonablemente Buena ● Regular





EDA Y MACHINE LEARNING

- Se fusionaron datasets en formato CSV relacionados con la calidad del aire.
- Se eliminaron valores nulos y se aplicaron procesos de limpieza de datos.
- Se entrenó un modelo de **clasificación** para predecir el **Índice de Calidad del Aire (ICA)** a partir de variables asociadas a contaminantes.
- Se desarrolló un modelo de **regresión** para estimar los niveles de **ozono** en función de datos climáticos.





RAG: Técnicas utilizadas

- Los documentos PDF fueron divididos en fragmentos (chunks) por página, y posteriormente cada fragmento fue indexado en Elasticsearch situado en una máquina virtual de azure para facilitar su búsqueda y recuperación contextual.
- Se utiliza **Flask** para construir una API REST que recibe preguntas del usuario.
- Se realiza una búsqueda semántica en **Elasticsearch** sobre un índice de documentos de calidad del aire.
- Los fragmentos relevantes se integran en un **prompt contextualizado**.
- La respuesta generada se limpia y se devuelve en **formato JSON**, incluyendo enlaces a los PDFs originales.
- En la interfaz web, se referencian los documentos PDF utilizados como fuente y se permite al usuario acceder directamente a su contenido completo para consulta contextual.



RAG

Elastic Search

Información en chUnks para el RAG

Busca informacion tecnica s Buscar

Resultados encontrados: 10

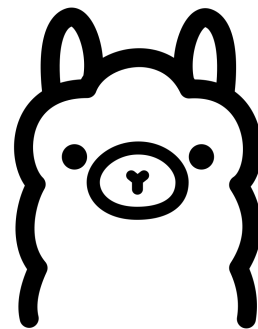
Resultados

- **Documento:** Informe evaluacion calidad aire España 2023.pdf
Página: 208
► **Contenido:** [Mostrar/ocultar](#)
- **Documento:** ECA 2023 INFORME COMPLETO.pdf
Página: 205
► **Contenido:** [Mostrar/ocultar](#)
- **Documento:** Informe evaluacion calidad aire España 2023.pdf
Página: 207
► **Contenido:** [Mostrar/ocultar](#)
- **Documento:** calidad-aire-estado-espanol-2023-informe.pdf
Página: 150
► **Contenido:** [Mostrar/ocultar](#)
- **Documento:** calidad-aire-estado-espanol-2023-informe.pdf
Página: 84
► **Contenido:** [Mostrar/ocultar](#)
- **Documento:** calidad-aire-estado-espanol-2023-informe.pdf
Página: 3

En este sistema **RAG (Retrieval-Augmented Generation)** se han indexado documentos relacionados con la calidad del aire en España a lo largo de los años. A través de la interfaz web, es posible realizar búsquedas para localizar los documentos que contienen información relevante en función de la pregunta del usuario.



Chatbot: Técnicas utilizadas



- Se utiliza **Flask** para construir una **API REST** que recibe preguntas del usuario.
- Se han creado funciones independientes para cada funcionalidad usando tools (temperatura, calidad del aire, búsqueda en Elasticsearch).
- En la búsqueda, se combinan búsquedas por palabras clave y vectoriales para mejorar precisión y relevancia.
- La función de temperatura y la de calidad de aire decide la fuente de datos según la fecha solicitada por el usuario: CSV para datos históricos y API para datos actuales.
- Se utiliza **Ollama**, instalada en el servidor **Oráculo**, para generar **embeddings** semánticos mediante un modelo de inteligencia artificial, incorporando control de errores para garantizar la robustez del proceso.



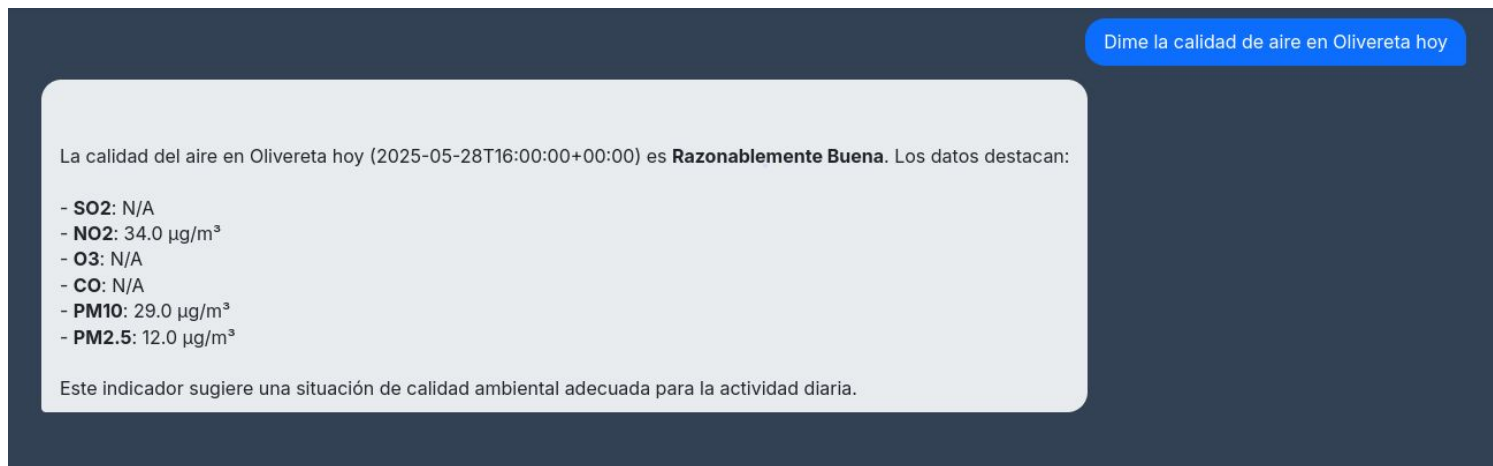
Chatbot: RAG

Busca informacion tecnica sobre catalunya

La información técnica sobre Catalunya incluye estudios de calidad del aire en regiones del país, iniciativas gubernamentales y datos ambientales relevantes. En Catalunya, se observan impactos ambientales significativos, como la reducción de contaminación en áreas del litoral y su relación con costes sanitarios y ambientales actuales. Estos estudios destacan la importancia de la mejora de la calidad del aire para la salud pública, reflejados en estudios como el de 2019 y el de 2022, que establecen límites legales para la utilización de combustibles altamente contaminantes. Además, se menciona la relevancia del impacto ambiental en la salud humana, incluyendo datos sobre mortales prematuras, enfermedades cardiovasculares y asma, que muestran cómo la calidad del aire afecta la salud pública.

En la captura se muestra una consulta al modelo **Ollama**, en la que se le solicita buscar en un servidor de **Elasticsearch** documentos relacionados con la pregunta del usuario. El modelo formula la búsqueda, accede al índice correspondiente y recupera los documentos más relevantes para construir una respuesta adecuada.

Chatbot: Calidad aire actual



En la captura se muestra una **consulta al modelo Ollama** sobre la **calidad del aire en Olivereta**. Para responder, el modelo realiza una **llamada a una API**, obtiene un **JSON con los datos ambientales** y **extrae los valores relevantes** para proporcionar una respuesta precisa y contextualizada.



Chatbot: Calidad aire histórico

Dime la calidad del aire de la universidad politécnica el día 2025-05-20

La calidad del aire para la Universidad Politécnica el día 2025-05-20 (2025-05-28T16:00:00+00:00) fue:

- SO2 (Partículas de óxido de sulfuro): 2.0
- NO2 (Partículas de nitrógeno de ozono): 6.0
- O3 (Ozone): 60.0
- CO (CO2): N/A
- PM10 (Partículas de metafórmica): 23.0
- PM2.5 (Partículas de microplaga): 10.0

El clima está considerado "Razonablemente Buena". Si necesitas más datos sobre otros días, por favor indica.

En la captura se muestra una consulta al modelo **Ollama** sobre la calidad del aire en la **Universidad Politécnica** en un día concreto. Para responder, el modelo busca en el csv, obtiene la fila con los datos ambientales de dicho día y extrae los valores relevantes para proporcionar una respuesta precisa y contextualizada.



Chatbot: Temperatura actual y historico

Cual es la temperatura en viveros hoy

La temperatura en la estación **ESTACI JARDINES DE VIVEROS** hoy (2025-05-28) es **24.4°C**. Las condiciones son:

- Humedad: 55.0%
- Viento: 1.7 m/s (dirección y coordenadas proporcionadas).

El tiempo actual es **2025-05-28T15:50:00+00:00**.

El día 2025-05-14 que temperatura hizo en la estación de viveros

La temperatura en la estación de Viveros el día 2025-05-14 fue **21.2°C**.